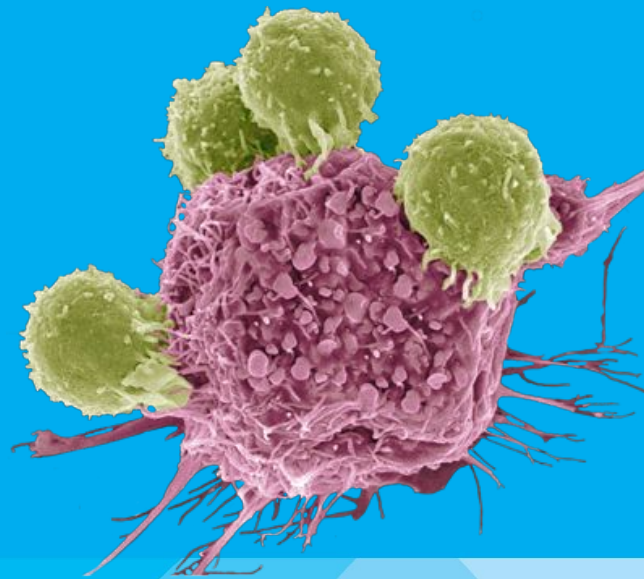# Machine Learning For Personalized Cancer Vaccines

Alex Rubinsteyn
February 9th, 2018
Data Science Salon Miami

OpenVax @ Mount Sinai

- Focus: **personalized cancer vaccines**
  - Machine learning for immunology
  - Cancer genomics
- Enthusiastically translational research
- Open source software: github.com/openvax
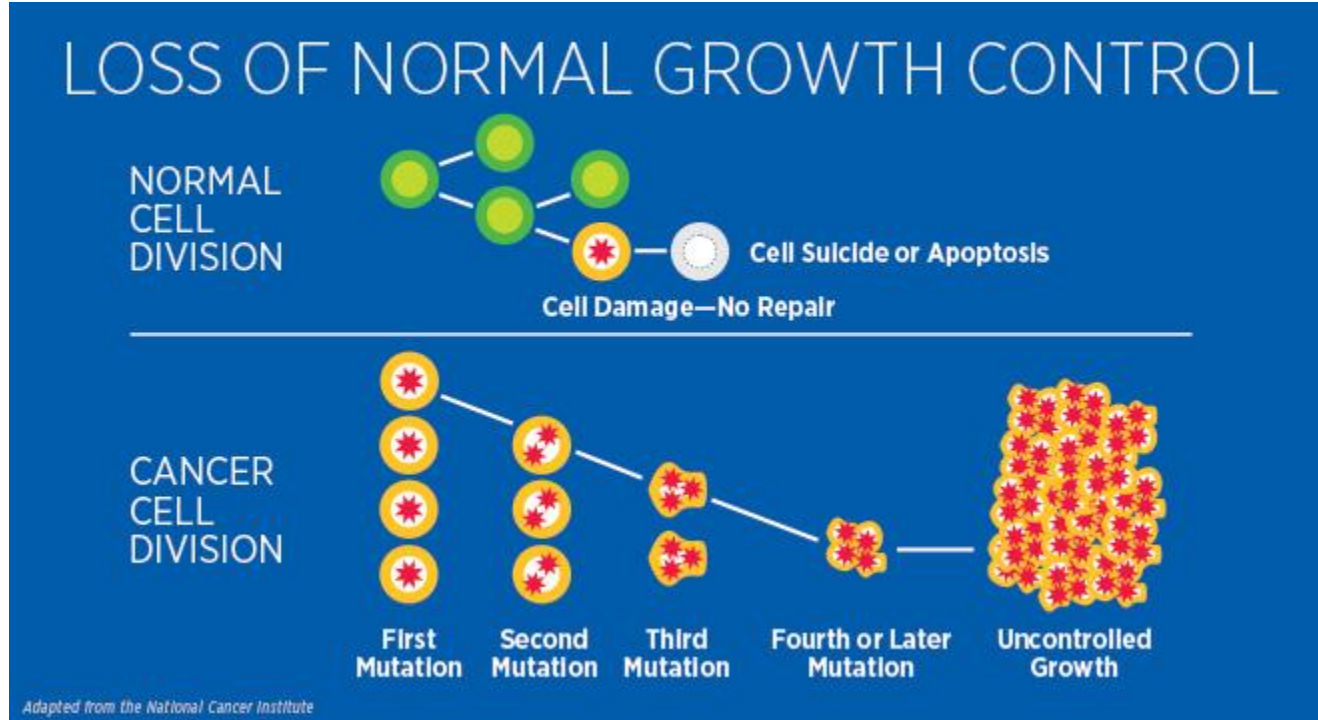- Website: www.openvax.org
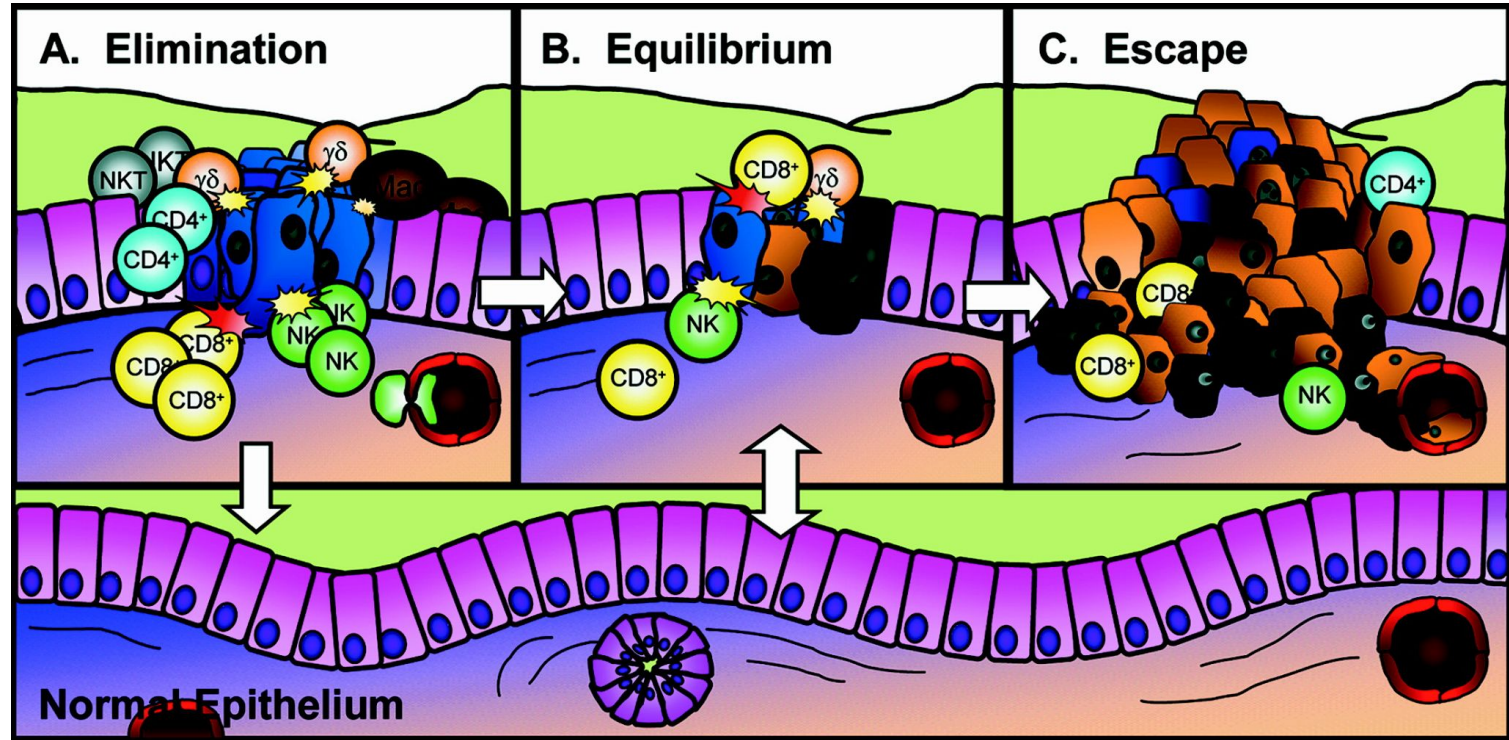
Icahn School
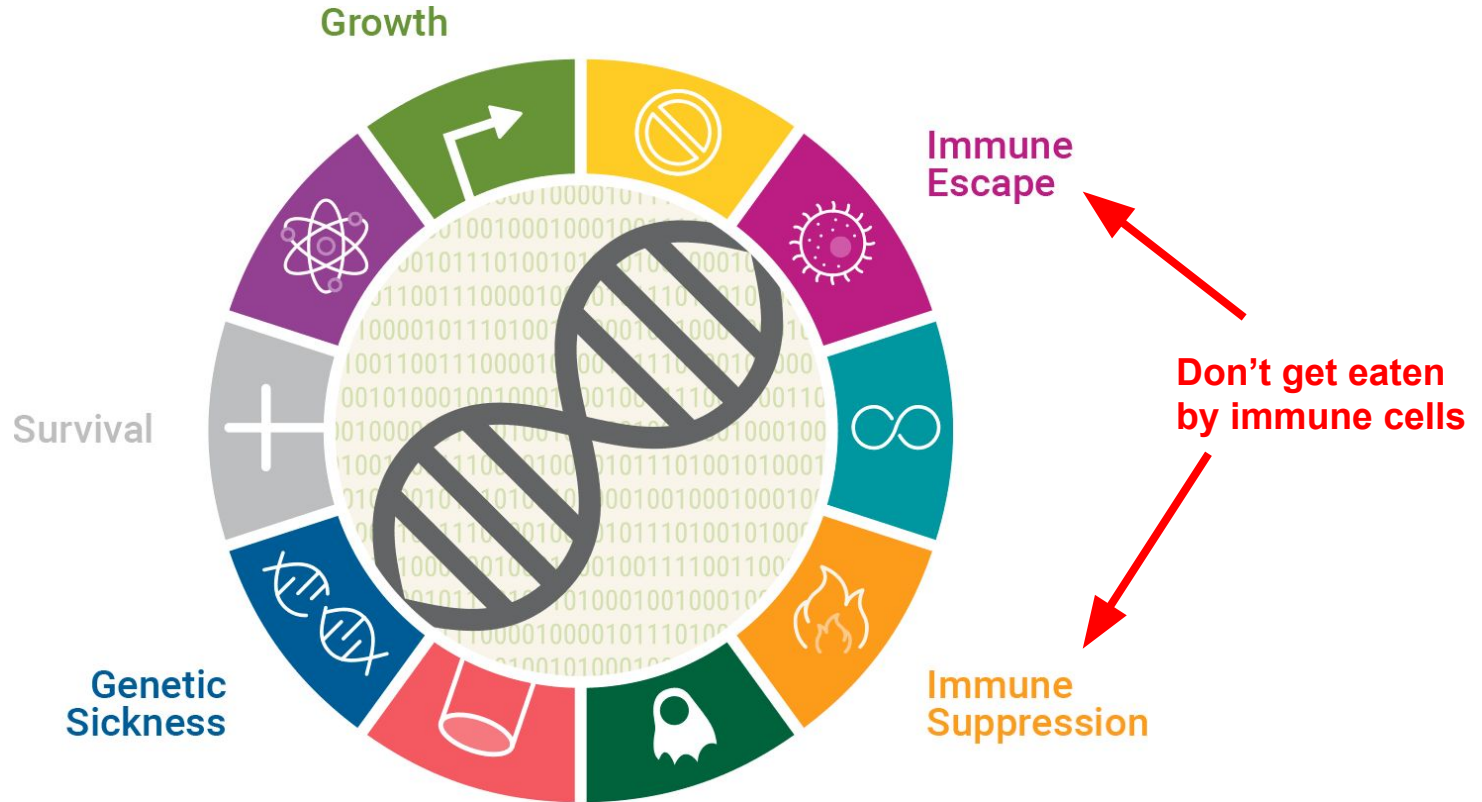of Medicine at
**Mount
Sinai**

Cancer Immunotherapy

What is cancer?

Immune system kills (most) cancer cells

Three E's of cancer immunity, Ian York (2007)

Immune avoidance a hallmark of cancer

Growth

Immune
Escape

Don't get eaten
by immune cells

Survival

Immune
Suppression

Hallmarks of
Cancer: The Next
Generation
(2011)

Genetic
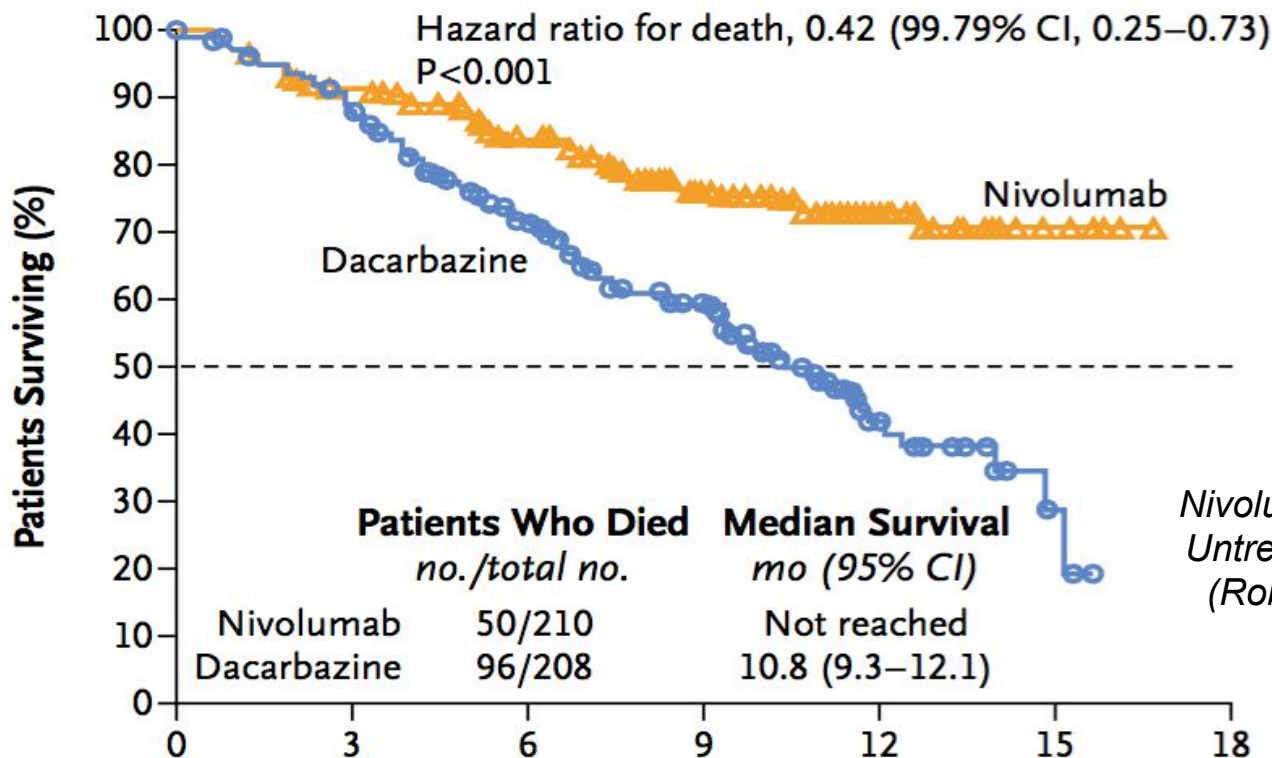Sickness

Cancer immunotherapy

- **Traditional treatments:** focus on killing cancer cells directly
- **Immunotherapy**: get the immune system to kill the cancer

- *Why is cancer spreading despite the immune system?*
  - Cancer cells inhibiting immune cells
    - Block the inhibitory signals!
  - Immune cells unable to recognize cancer as non-self
    - Teach the immune system what to kill

Immunotherapy vs. Chemotherapy

Overall Survival

Hazard ratio for death, 0.42 (99.79% CI, 0.25–0.73)
P<0.001

Nivolumab

Dacarbazine

Nivolumab in Previously Untreated Melanoma… (Robert NEJM 2015)

	Patients Who Died no./total no.	Median Survival mo (95% CI)
Nivolumab	50/210	Not reached
Dacarbazine	96/208	10.8 (9.3–12.1)

Therapeutic Cancer Vaccines

Unpacking "therapeutic cancer vaccine"

- **Therapeutic**
  - Treating established disease (not preventative)
- **Cancer Vaccine**
  - Teach the immune system to kill cancer

What's in a therapeutic cancer vaccine?

- Tumor antigen
  - What should immune system look for?
- Adjuvant
  - Something the immune system already responds to as dangerous
  - Examples: double-stranded RNA, mineral oil, dead bacteria
- *Objective*: get the immune system to learn that the antigen is bad and cells which have it should be killed

Personalized therapeutic cancer vaccines

- **Personalized**
  - Made from scratch for each patient
  - Requires profiling of patients & their tumors
- **Therapeutic**
  - Treating established disease (not preventative)
- **Cancer Vaccine**
 - Teach the immune system to kill cancer

Choosing what goes in the vaccine

- Sequence patient & tumor DNA
  - Identify tumor mutations
- Sequence tumor RNA
  - Which mutations are being produced into proteins?
- Predict which mutations can be seen by immune system

Computational genomics tools for dissecting tumour–immune cell interactions
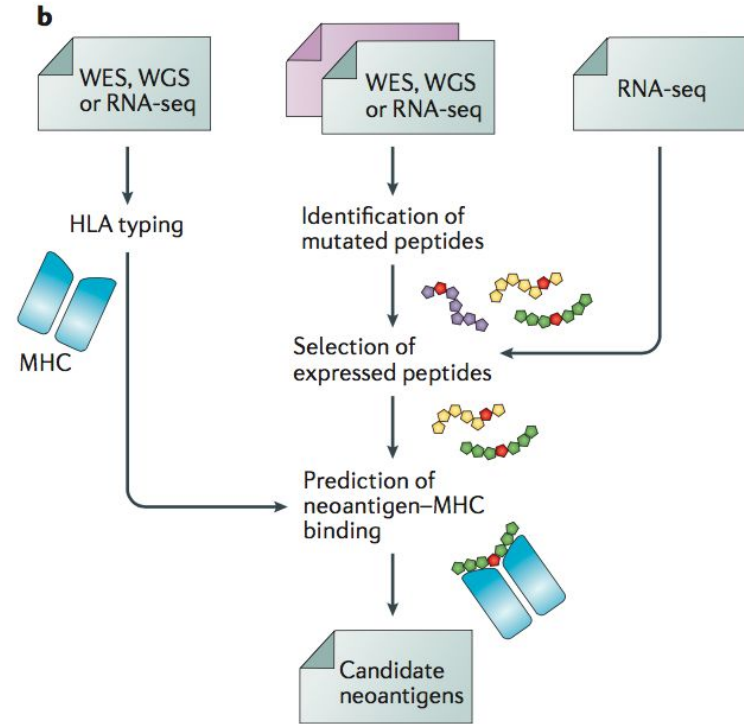
Does it work in mice?

- Details:
  - BALB/c mice
  - CT26 colon cell line
  - mRNA vaccine
- Two groups of 5 mutations
  - Individual mutatations don't work
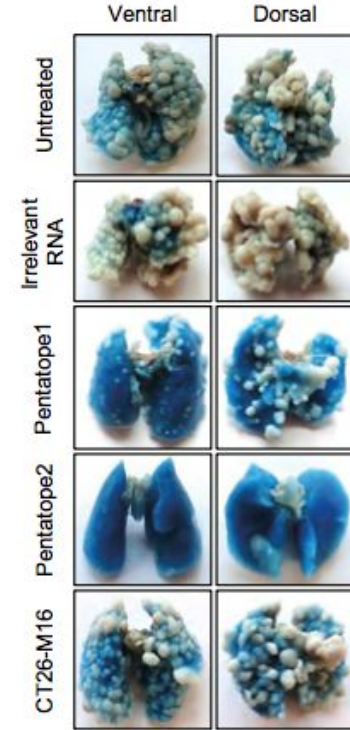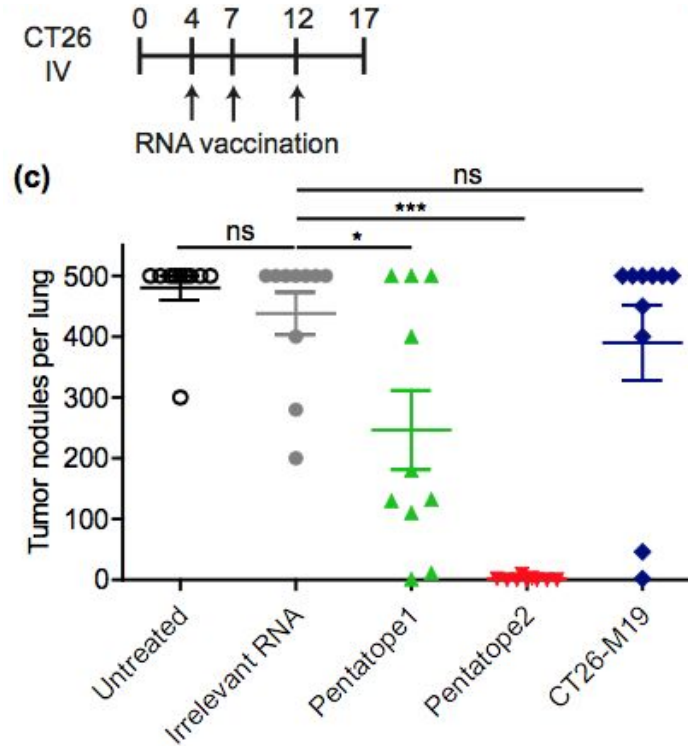
Mutated neo-antigens as targets for individualized cancer immunotherapy (Figure 3.18), Vormehr (2016)

Clinical trial at Dana Farber

An immunogenic personal neoantigen vaccine for patients with melanoma

Patrick A. Ott[1,2,3]*, Zhuting Hu[1]*, Derin B. Keskin[1,3,4], Sachet A. Shukla[1,4], Jing Sun[1], David J. Bozym[1], Wandi Zhang[1], Adrienne Luoma[5], Anita Giobbie-Hurder[6], Lauren Peter[7,8], Christina Chen[1], Oriol Olive[1], Todd A. Carter[4], Shuqiang Li[4], David J. Lieb[4], Thomas Eisenhaure[4], Evisa Gjini[9], Jonathan Stevens[10], William J. Lane[10], Indu Javeri[11], Kaliappanadar Nellaiappan[11], Andres M. Salazar[12], Heather Daley[1], Michael Seaman[7], Elizabeth I. Buchbinder[1,2,3], Charles H. Yoon[3,13], Maegan Harden[4], Niall Lennon[4], Stacey Gabriel[4], Scott J. Rodig[9,10], Dan H. Barouch[3,7,8], Jon C. Aster[3,10], Gad Getz[3,4,14], Kai Wucherpfennig[3,5], Donna Neuberg[6], Jerome Ritz[1,2,3], Eric S. Lander[3,4], Edward F. Fritsch[1,4]†, Nir Hacohen[3,4,15] & Catherine J. Wu[1,2,3,4]

- 6 (stage III & IV) melanoma patients
- Up to 20 mutations per vaccine
- Adjuvant: Poly-ICLC *(synthetic double-stranded RNA)*

Dana Farber Trial: Tumor Control
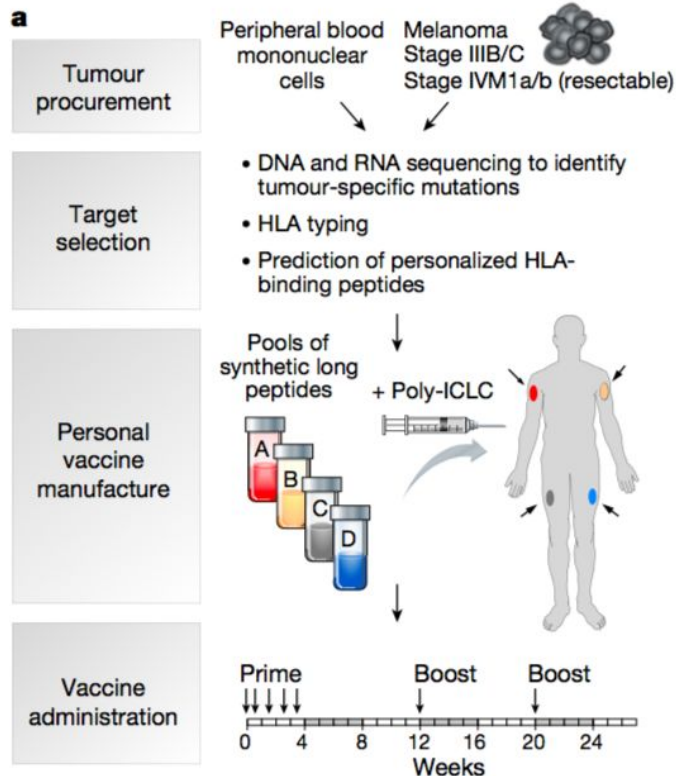
Of six vaccinated patients, four had no recurrence at 25 months after vaccination, while two with recurrent disease were subsequently treated with anti-PD-1 (anti-programmed cell death-1) therapy and experienced complete tumour regression, with expansion of the repertoire of neoantigen-specific T cells.
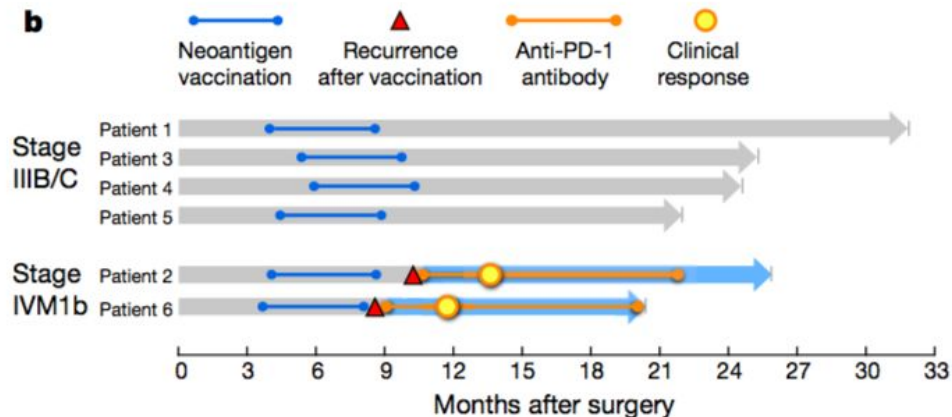
Machine Learning for Predicting Immune Responses

Where do ML models fit in?

- Vaccines typically target 5-20 mutations
- Depending on the cancer types, we might have hundreds of expressed mutations
- Immune predictions pick the winners

Starting material: Tumor DNA and RNA

Exome sequencing: Nonsynonymous variants: 92

RNA-sequencing: Transcribed variants: 39

Epitope prediction: Targets: 17

Low Mutation Burden in Ovarian Cancer May Limit the Utility of Neoantigen-Targeted Vaccines (2016)
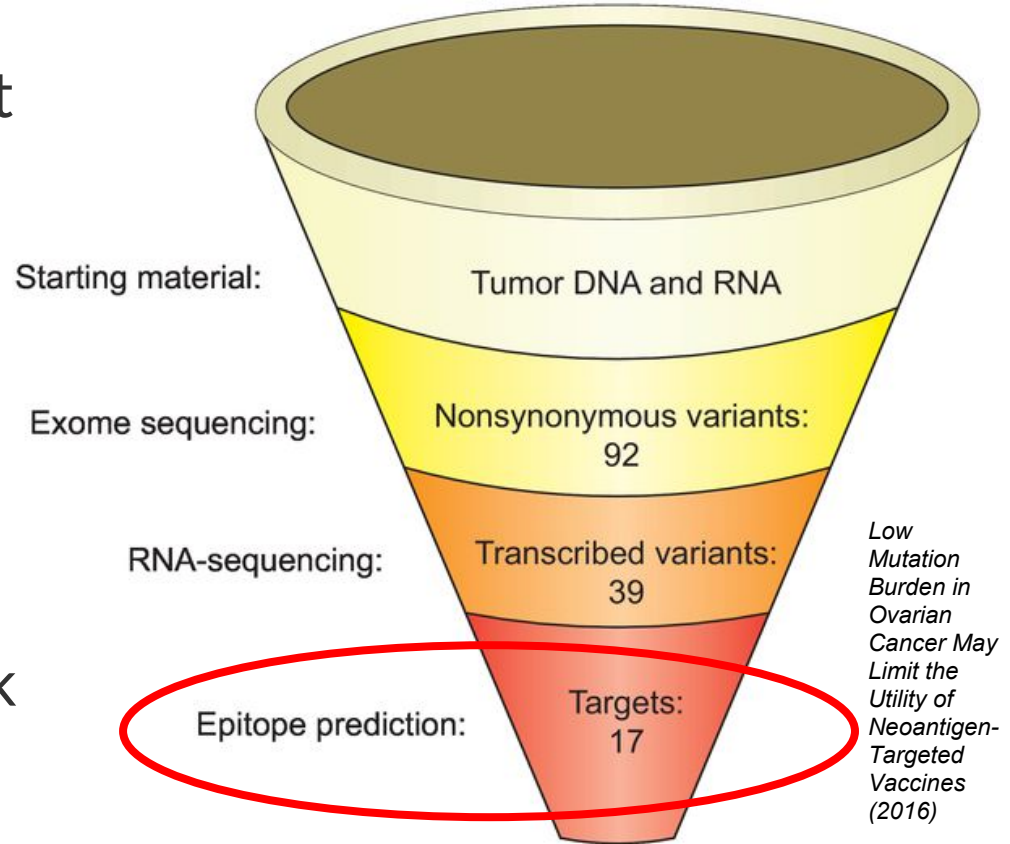
Quick intro to T cells

- Proteins cleaved into peptides
- Some peptides loaded on MHC
- Peptide/MHC complexes presented on cell surface
- T cells look at peptide/MHC complexes
- Abnormal displayed peptides lead to a cytotoxic T cell response

Yewdell, J.W., Reits, E. & Neefjes, J., 2003. *Nature Reviews Immunology*
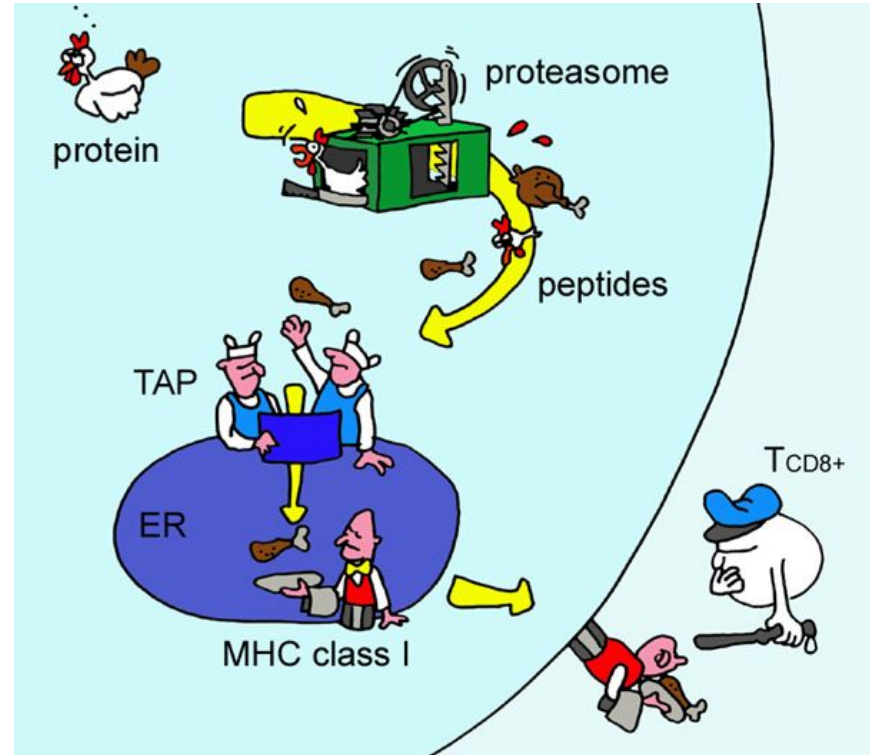
MHC Binding Prediction

- Thousands of MHC alleles in human population

- Each allele capable of binding a distinct set of peptides

- *Objective:* Predict whether an MHC allele will bind a given peptide

Holland, C.J., Cole, D.K. & Godkin, A., 2013. *Frontiers in Immunology*
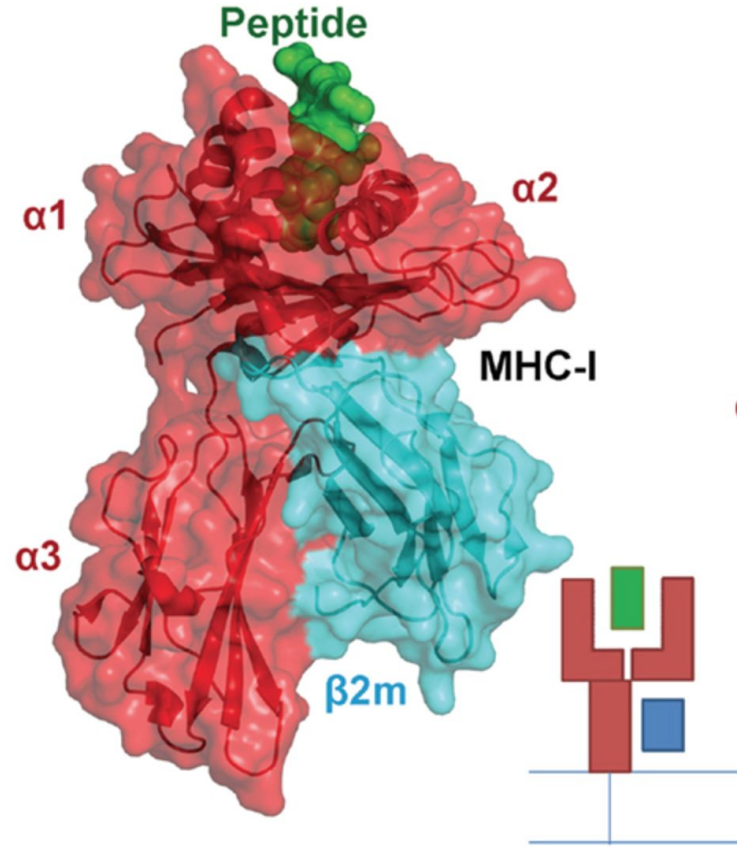
Immune Epitope Database (IEDB)

- Public dataset of immunology data

- Includes >200,000 *in vitro* binding affinity measurements of purified MHC/peptides

- Core training data for MHC ligand prediction tools

Summary Metrics	
Peptidic Epitopes	378,287
Non-Peptidic Epitopes	2,543
T Cell Assays	323,210
B Cell Assays	395,472
MHC Ligand Assays	782,035
Epitope Source Organisms	3,627
Restricting MHC Alleles	751
References	18,714

MHC Binding Data

MHC Allele Name	Epitope Description	Assay Quantitative measurement
HLA-A*11:01	ADLVGFLLLK	20.300000
HLA-A*03:01	ALAETSYVK	35.500000
HLA-A*11:01	ALAETSYVK	16.400000
HLA-A*02:01	ALAETSYVKV	333.300000
HLA-A*02:01	ALGLVCVQA	333.300000
HLA-A*02:01	ALGLVCVQM	5000.000000
HLA-A*02:01	ALREEEGV	238.100000
HLA-A*03:01	DLVGFLLLK	2750.000000
HLA-A*03:01	DLVGFLLLKY	1594.200000

T Cell Response Data

antigen	assay	mhc	organism	peptide	response
Assembly protein G7	IFNg release	HLA-A*26:01	Vaccinia virus	ESKAKQLCY	Negative
Protein A40	IFNg release	HLA-B*40:01	Vaccinia virus	IETPNELSF	Negative
Major core protein 4a precursor	IFNg release	HLA-A*03:01	Vaccinia virus	VTNLISETLK	Negative
NaN	CCL4/MIP-1b release	HLA-A*02:01	Homo sapiens	ILAKFLHWL	Positive
Protein F16	IFNg release	HLA-A*33:03	Vaccinia virus	RFVNKLKMYK	Negative

Linear models perform reasonably well

- One-hot encoding of amino acids at each position
- **I**gnore dependencies between positions
- AUC ~= 0.85-0.9

	HLA A*0201								
	1	2	3	4	5	6	7	8	9
A	-0.3	0.8	-0.3	-0.3	-0.2	-0.3	0.0	0.0	-0.9
C	0.2	0.9	0.0	0.3	-0.5	-0.1	0.1	0.2	0.4
D	0.8	0.9	-0.4	-0.3	0.3	0.2	0.4	0.3	0.6
E	0.6	-0.4	0.7	-0.2	0.1	-0.4	-0.2	-0.2	-0.5
F	-1.3	0.5	-0.5	0.1	-0.1	0.0	-0.3	-0.4	-0.8
G	-0.2	0.1	0.3	-0.1	0.0	0.4	0.3	-0.1	0.2
H	1.1	0.9	-0.1	0.4	0.1	0.2	0.0	0.2	0.8
I	-0.4	-0.7	-0.4	0.1	-0.1	-0.4	-0.5	0.5	-1.4
K	-0.3	0.0	1.1	0.1	0.1	0.6	0.9	0.2	0.9
L	0.0	-1.9	-0.4	-0.2	0.0	-0.2	0.0	-0.1	-1.1
M	-0.7	-1.2	-0.7	0.2	-0.6	0.0	0.0	0.0	-0.8
N	-0.1	0.3	0.1	-0.3	-0.1	-0.3	0.0	0.2	0.7
P	1.2	0.5	0.6	-0.3	0.4	0.0	-0.4	-0.5	0.7
Q	0.4	-1.1	0.0	-0.1	0.4	-0.2	-0.3	0.2	0.7
R	-0.2	0.9	1.0	0.3	0.1	0.4	0.7	0.0	0.9
S	-0.3	0.1	0.1	-0.4	0.1	0.3	-0.2	-0.1	0.2
T	-0.2	-0.5	0.1	0.4	0.1	-0.5	0.2	0.0	-0.1
V	-0.1	-0.9	-0.1	0.2	0.0	-0.3	0.1	0.1	-1.9
W	0.0	0.7	-0.5	-0.2	-0.1	0.2	-0.3	-0.1	0.4
Y	-0.3	0.2	-0.6	0.2	0.0	0.4	-0.4	-0.3	0.8

Bjoern Peters

Neural networks do better: NetMHCpan

- Standard tool to predict peptide/MHC binding affinity
- Inputs: peptide sequence and binding groove residues of MHC
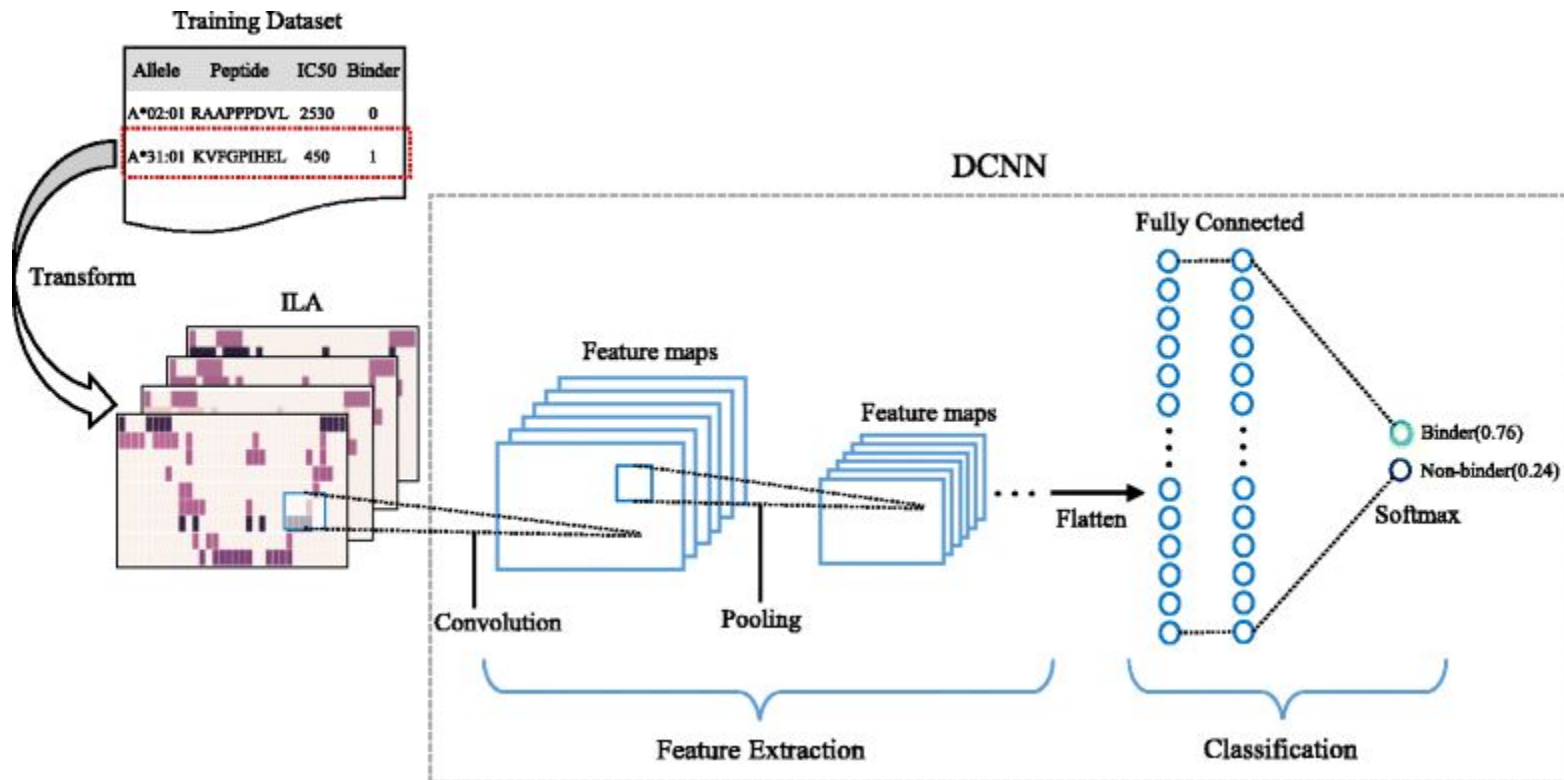

SOFTWARE | **Open Access**

NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets

Morten Nielsen[1,2*] and Massimo Andreatta[1]

Deep learning meets immunology

Deep convolutional neural networks for pan-specific peptide-MHC class I binding prediction

Python Tools for Immunology ML

- **Pandas**
  - Load and filter messy immunology data
- **Seaborn**
  - Visualization
- **Scikit-learn**
  - Model selection, metrics, shallow (linear & tree) models
- **Keras or PyTorch**
  - Deep learning / neural networks

Startup Landscape

Funding for Personalized Cancer Vaccines

moderna™   **$1.2B**

BIONTECH   **$270M**

gritstone ONCOLOGY   **$195M**

NEON THERAPEUTICS   **$161M**

+ ~20 more companies

Thanks!